

FIG. 1

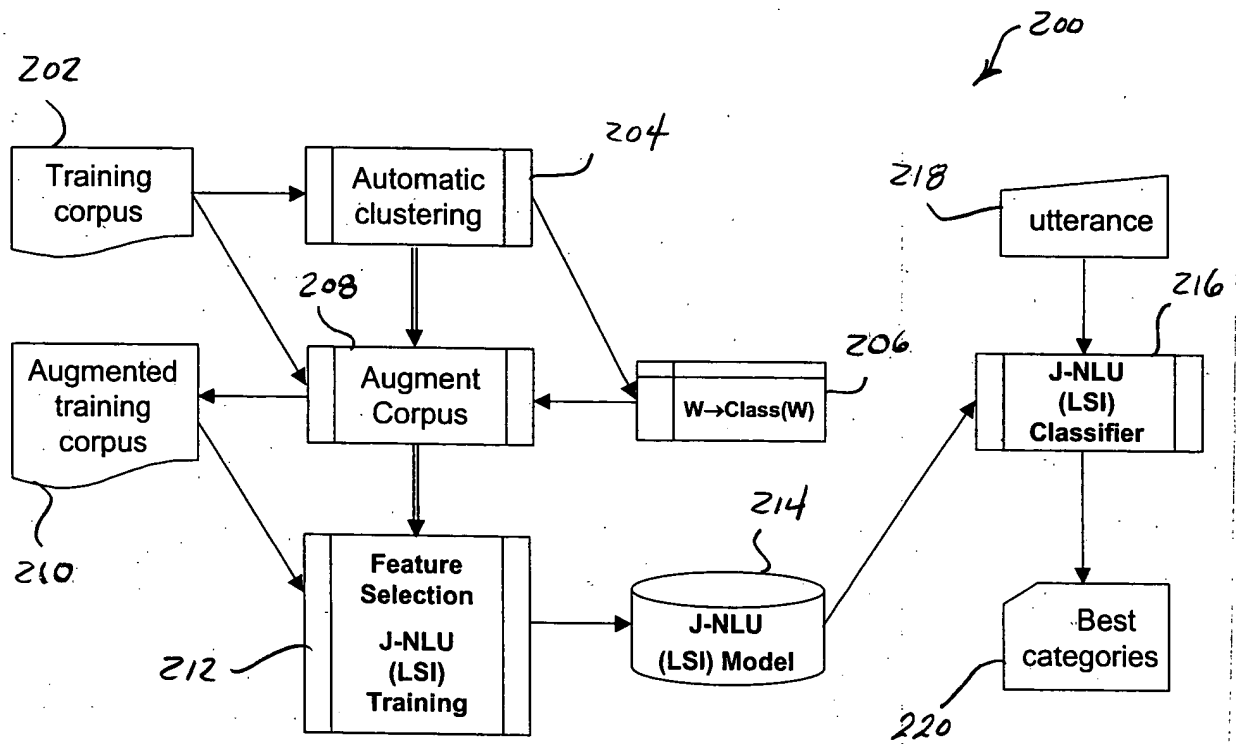


FIG. 2

3/5

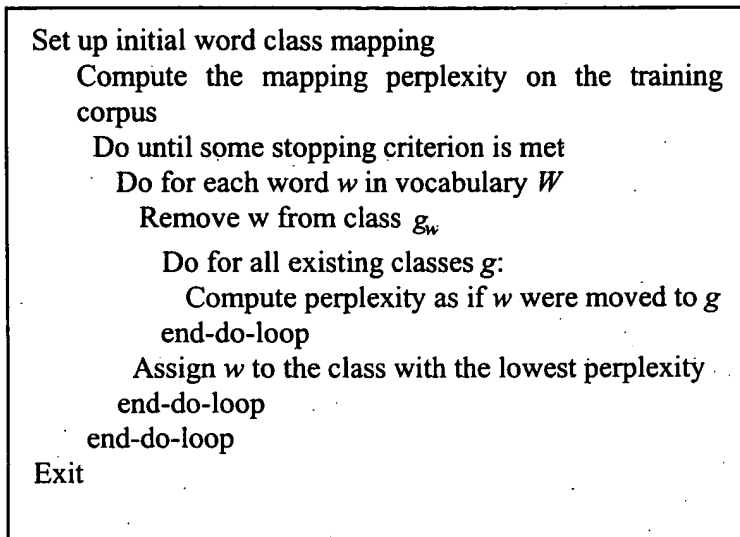


FIG. 3

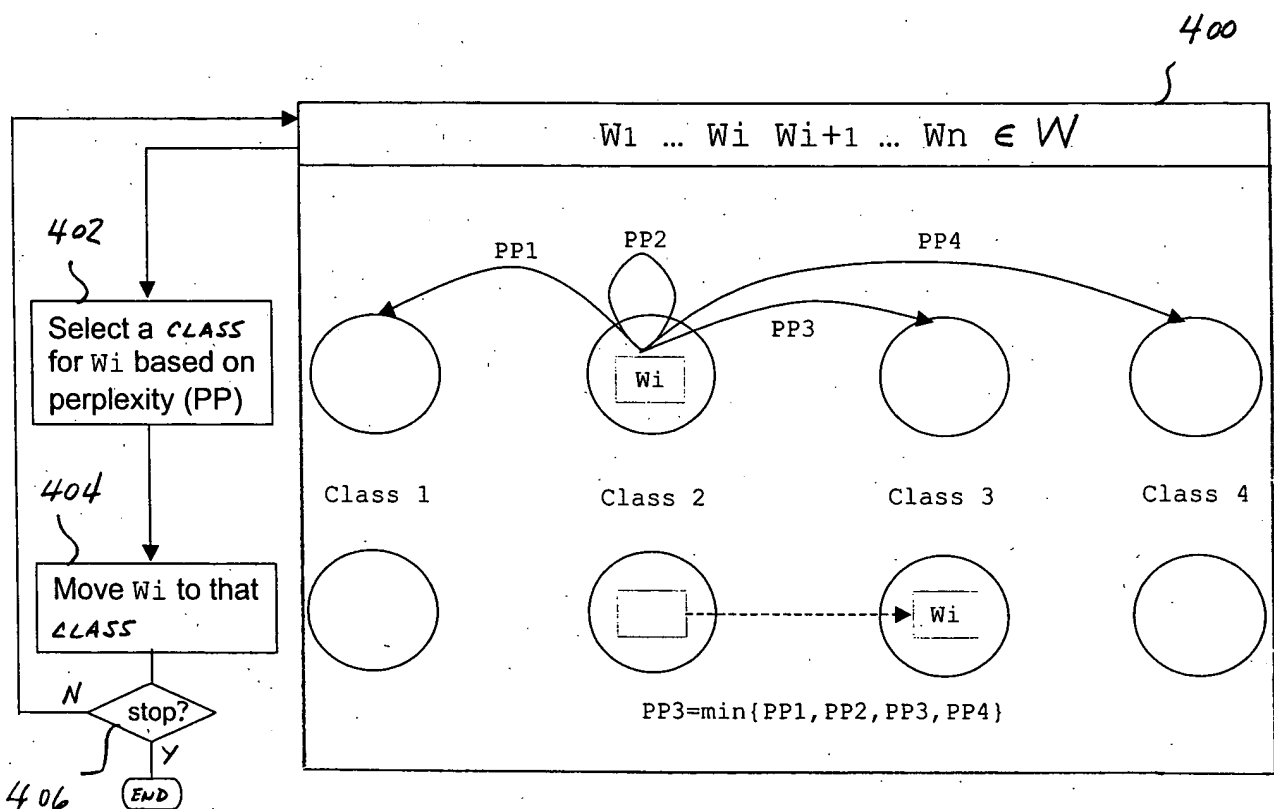
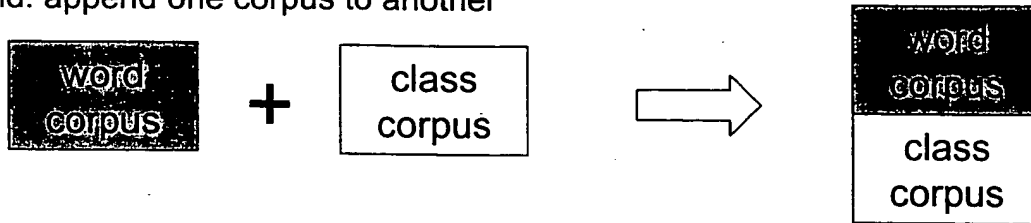


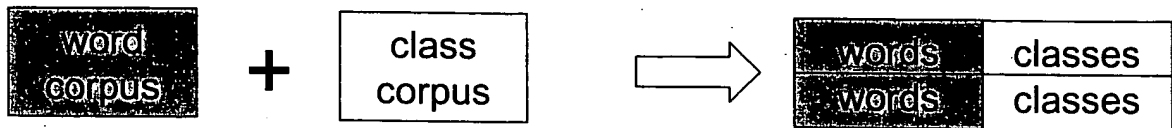
FIG. 4

4/5

Append: append one corpus to another



Join: join the utterances in two corpora



Interleave: insert each class after its word

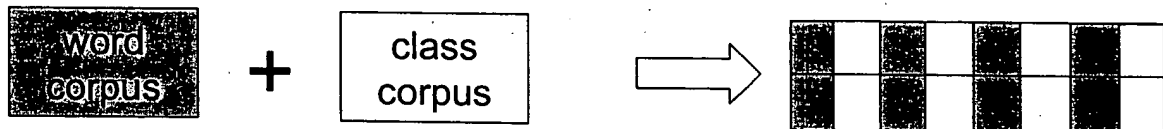


FIG. 5

Given a parameter p :

1. Calculate the IG value of each term
2. Sort the terms by their IG values into descending order
3. Set the threshold t to the IG value at top p percentile of sorted terms
4. Select terms with IG value $\geq t$

FIG. 6

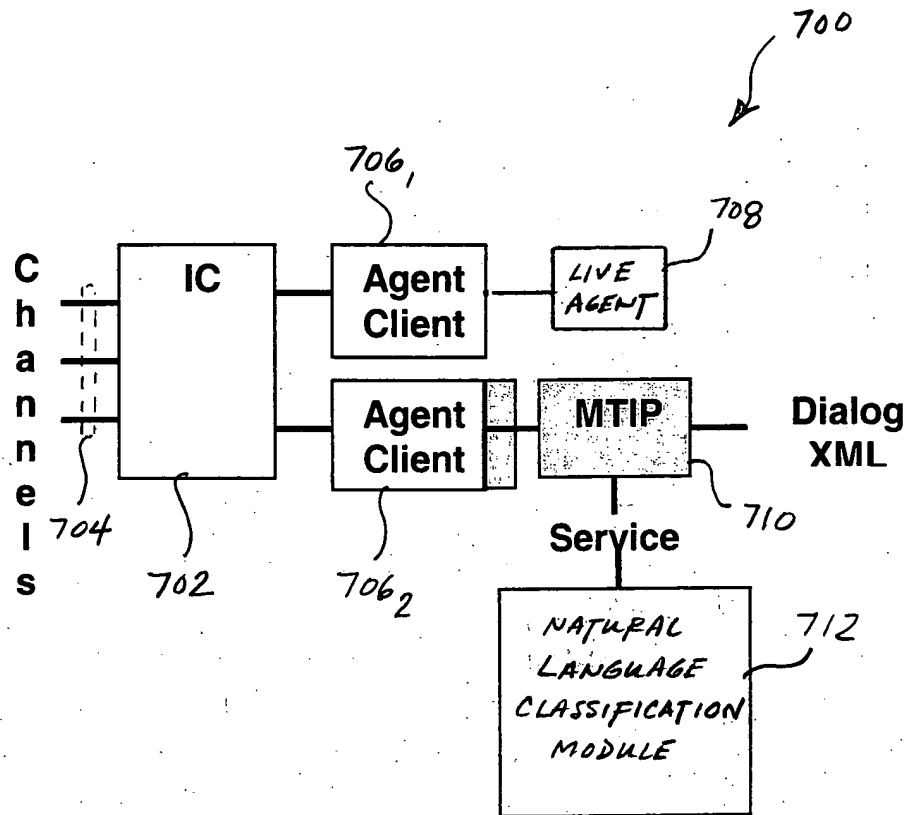


FIG. 7